

An information-carrying and knowledge-producing molecular machine. A Monte-Carlo simulation

Christoph Kuhn

Received: 10 February 2011 / Accepted: 29 March 2011 / Published online: 11 May 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The concept called *Knowledge* is a measure of the quality of genetically transferred information. Its usefulness is demonstrated quantitatively in a Monte-Carlo simulation on critical steps in a origin of life model. The model describes the origin of a bio-like genetic apparatus by a long sequence of physical-chemical steps: it starts with the presence of a self-replicating oligomer and a specifically structured environment in time and space that allow for the formation of aggregates such as assembler-hairpins-devices and, at a later stage, an assembler-hairpins-enzyme device—a first translation machine.

Keywords Assembler-hairpins-enzyme device · Emergence and storage of information · Genetic apparatus · *Knowledge* · Monte-Carlo · Origin of life · Structured time-space environment

Introduction

Modeling the origin of life: postulates on initial conditions

This paper is based on the concept that living individuals—distinct aggregates of interlocking molecules—are a form of matter that carries information to be reproduced and to evolve in their given environment into forms of increasing complexity and intricacy [1–7].

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1081-9) contains supplementary material, which is available to authorized users.

C. Kuhn (✉)
Biomedical Optics Research Laboratory, Clinic of Neonatology,
University Hospital Zürich,
Frauenklinikstrasse 10,
8091 Zurich, Switzerland
e-mail: c-k@gmx.ch

A step-by-step Darwinian process appears instantaneous when the first entity emerges by chance (a de-novo oligomer that is capable of replication), survives as a species (population of individuals of a certain kind) and then evolves by continued reproduction with variation and selection in the particular environment driving the process. The initial step, which has been called "origin of life", requires a basic initial form for the building blocks as well as basic initial conditions for the environment (Supplementary Fig. 1). Such a step-by-step Darwinian process, continuous and sustained, leads first to aggregates of interlocking molecules, and then to the bio-genetic apparatus.

These aggregates reach increasing independence from the afore-mentioned highly specific initial conditions by gradually populating a diversified area, as follows: (1) populating regions of compartments with increasing size of pores, (2) evolving devices that produce envelopes as innate compartments, (3) evolving metabolism (increasing intricacy of the living machinery).

Modeling origin of life: the RNA world

We consider the case where only one type of complementary R-monomers, R_1 and R_2 , is present, and we assume that de-novo oligomers emerge and are capable of replicating (Supplementary Fig. 2a). The initial sequence is random and, due to errors in the copying process, there is variation of this sequence. If by chance a strand is found in the population that has a sequence such that it folds onto itself to form a hairpin (Supplementary Fig. 2c), the hairpin will be selected because it is more protected against hydrolysis and thus has a higher probability of being selected. The hairpin finds an open strand to bind by the loop by complementary and anti-parallel means. If the number of adjacent hairpins equals or exceeds three, such an aggregate will be selected because it is, again, more protected against hydrolysis, and thus has a higher probability of being selected. A reading

frame (Supplementary Fig. 2d) exists if all hairpins in the population have loops of one kind. All hairpins bound to the assembler are then adjacent (Supplementary Fig. 2e).

The early process (RNA-world [8–11]) comes to an end when replication reaches a certain precision. This apparent dead-end is overcome by the emergence of the first translation apparatus with an enzyme E_1 as its product (i.e., the assembler-hairpins-enzyme device, the HAE_1 -device and its nonsensical replica, the HAE_0 -device as explained below).

Modeling the origin of life: the RNA–protein world

We introduce in our model two a-monomers, a_1 and a_2 , (which may be glycine and alanine). Occasionally, by errors in the copying process, the two R-monomers at the open end (Supplementary Fig. 3a) of the hairpin are no longer complementary. The hairpin then carries an a-monomer such that R_1 – R_1 carries a_1 and R_2 – R_2 carries a_2 (Supplementary Fig. 3a). If all hairpins on an assembler carry a-monomers, these can then oligomerize. The product, an a-oligomer with random sequence, has no enzymatic power but functions as an agglutinate-forming envelope E_0 (Supplementary Fig. 3b).

The HAE_1 -device is a translation apparatus (Supplementary Fig. 3c), and emerges as a by-product. It is selected only if (1) there is a code on the (+)-assembler for this specific a-oligomer, (2) this code is translated by the hairpins, and (3) the specific a-oligomer is an enzyme (enzyme E_1) that increases the precision of the R-replication (replicase). The (–)-assembler constitutes the anti-parallel complementary copy of an (+)-assembler, and its product therefore has no sense (HAE_0 -device).

Increasing contamination coupled with increasing complexity of the evolving system (i.e., by accumulation of HAE_0 -devices with an increasing number of different HAE_i -devices) again leads to another barrier. This barrier is overcome by a fundamental change in the machinery into a primordial form of the translation apparatus (DNA–RNA–protein world [12–16]). The emergence of a bio-genetic apparatus [7] then paves the way for the “explosion of life” [17–19].

Computer implementation and simulation

Supplementary Fig. 4 presents the computer implementation [20, 21] in a flow chart overview format. The top part shows the construction phase, where the aggregates are formed. According to the fitness of the formed devices, parameters such as survival chance, replication probability and the probability of occurrence of an error in the replication, are assigned to these entities. Possible aggregates are the hairpin (Supplementary Fig. 2c), the HA-device (Supplementary Fig. 2e) and, after emergence of new kinds of monomers (a-monomers) that attach to the open ends of hairpins, the

HAE_0 - (Supplementary Fig. 3a) and the HAE_1 - (Supplementary Fig. 3c) devices. In the selection phase (middle of Supplementary Fig. 4), only a fraction of the population survives (corresponding to half of the total number of strands). In the multiplication phase (bottom of Supplementary Fig. 4), the aggregates dissociate and the strands are copied by chance until the total number of strands is replenished. Two typical simulations starting with a single strand of random sequence are shown in Supplementary Fig. 5a,b.

Quality of information *Knowledge* demonstrated by a Monte Carlo study

Defining quantities that measure *Information* and *Knowledge* elucidates some principle aspects of the origin of life. Genetic information is stored in one entity and transferred from one generation to the next as a distinct sequence of monomers in a strand. It is measured in bits [22].

“The quality of genetic information” denotes that the genetic information has the property to instruct the formation of an entity that behaves as if it had knowledge, that is, it behaves as if it would know how to survive and to multiply in its given environment.

Knowledge, K , is related to the effort required to develop a certain degree of functionality. Let us consider the total information $I(g)$ (number of bits at generation g) that has to be discarded by eliminating unfit individuals along a singlet out trial of a Darwinian evolution that takes place within a population of entities from the initial replicating oligomer at generation 0, continued by many step-by-step Darwinian processes, until the given degree of functionality is reached at generation g . The average g_{bt} of the repeated process (Monte Carlo method) levels out random fluctuation in g . The effort required to develop the given degree of functionality is measured by the number $K=I(g_{bt})$ called *Knowledge* K [3, 7, 23].

Beginning with a situation in which the emergence of a first replicating strand is possible, it takes a huge number of trials (i.e., discarding much information) until a replicable strand actually appears (*Knowledge* K growing suddenly from zero to a value K_0 given by the sum of discarded bits (see schematic illustration in Supplementary Fig. 6a, b). Then, *Knowledge* K will stepwise increase, being constant (reflecting the gradual adaptation of the evolving forms to a given region by refinements of their organizational structure) until a major change in the organizational structure appears with a stepwise increase in *Knowledge* K (break-through process leading to the colonization of a new region). Supplementary Fig. 7a (analyzing the emergence of hairpins) and Supplementary Fig. 7b (showing *Knowledge* K of the evolution from replicating strands to the

HAE₁-device) are evaluated from the statistics of the Monte Carlo study. Building up the HA-device with its reading frame requires many generations, whereas the HAE₁-device emerges shortly after a-monomers appear that are able to link to the hairpin's open end.

Conclusions

Critical steps in a model on the origin of a bio-like genetic apparatus have been demonstrated by a computer simulation and analyzed by Monte-Carlo method. Establishing a reading frame, from a state of randomly formed aggregates of hairpins to a state of a well-ordered HA-device required many generations of a step-by-step Darwinian process. Modeling the origin of a bio-like genetic apparatus—from the emergence, de-novo, of a self-replicating oligomer and a specifically structured environment in time and space that allows the formation of aggregates such as assembler-hairpins-devices and continuing evolution to increasing complexity, which ultimately leads to the assembler-hairpins-enzyme device—is important to understand the principles of the origin of life and to stimulate testing the origin of life model by experimental efforts. The concept of *Knowledge* is therefore a useful measure for the quality of information along Darwinian evolution.

Acknowledgments I would like to express my gratitude to Hans Kuhn for his continued and sustained encouragement and Evelyn Bozzi for kindly copy editing the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Kuhn H (1972) Self-organization of molecular systems and evolution of the genetic apparatus. *Angew Chem Int Ed Engl* 11:798–820
- Kuhn H (1976) Model consideration for the origin of life. Environmental structure as stimulus for the evolution of chemical systems. *Naturwissenschaften* 63:68–80
- Kuhn H, Waser J (1983) Self-organization of matter and early evolution of life. In: Hoppe W, Lohmann W, Markl H, Ziegler H (eds) *Biophysics*. Springer, Berlin, pp 830–874
- Kuhn H, Waser J (1994) A model of the origin of life and perspectives in supramolecular engineering. In: Behr JP (ed) *Lock-and-key principle*. Wiley, Chichester, pp 247–306
- Kuhn H, Kuhn C (2003) Diversified world: Drive of life's origin?! *Angew Chem Int Ed* 42:262–266
- Kuhn H (2008) Origin of life—symmetry breaking in the universe: emergence of homochirality. *Curr Opin Colloid Interface Sci* 13:3–11
- Kuhn H (2010) Is the transition from chemistry to biology a mystery? *J Syst Chem* 1:3
- Orgel LE (1973) *The origins of life: molecules and natural selection*. Wiley, New York
- Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* 39:99–123
- Dworkin JP, Lazcano A, Miller SL (2003) The roads to and from the RNA-world. *J Theor Biol* 222:127–134
- Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459:239–242
- Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* 261:1411–1418
- Keefe AD, Szostak JW (2001) Functional proteins from a random sequence library. *Nature* 410:715–718
- Seelig B, Szostak JW (2007) Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* 448:828–831
- Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. *Nature* 409:387–390
- Shevack A, Gewitz HS, Hennemann B, Yonath A, Wittmann HG (1985) Characterization and crystallization of ribosomal particles from *Halobacterium marismortui*. *FEBS Lett* 184:68–71
- Crick FHC (1958) On protein synthesis. *Symp Soc Exp Biol* XII:139–163
- Crick FHC (1970) Central dogma of molecular biology. *Nature* 227:561–563
- Dawkins R (2004) *Ancestor's tale. A pilgrimage to the dawn of life*. Weidenfeld & Nicolson, London
- Kuhn C (2001) Computer-modelling origin of a simple genetic apparatus. *Proc Natl Acad Sci USA* 98:8620–8626
- Kuhn C (2005) A computer-glimpse of the origin of life. *J Biol Phys* 31 C:571–585
- Shannon CE, Weaver W (1949) *The mathematical theory of communication*. University of Illinois Press, Urbana, Illinois
- Kuhn H (1988) Origin of life and physics: diversified microstructure inducement to form information-carrying and knowledge-accumulating systems. *IBM J Res Dev* 32:37–46